DOCUMENT RESUME

ED 295 995 TM 011 851

AUTHOR Carlson, Sybil B.
TITLE Writing Samples Viewed from Different Perspectives: An Approach to Validity.
PUB DATE Apr 88
NOTE 41p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Processes; College Students; Construct Validity; *Discourse Analysis; Higher Education; *Holistic Evaluation; *Scoring; *Validity; Writing (Composition); *Writing Evaluation
IDENTIFIERS Graduate Record Examinations

ABSTRACT

The objective description and identification of variables that meaningfully distinguish reasoning skills couched in written discourse was studied, comparing scores obtained from different perspectives on the same writing samples. A total of 406 writing samples on 2 topics by 203 students who had taken the Graduate Record Examinations, mostly non-native speakers of English, was reviewed by the Writer's Workbench computer analysis and four different schemes: (1) the Purves-Soter scheme; (2) the Moss scheme; (3) the Reid holistic scoring scheme; and (4) another holistic scheme developed for a previous study. The reliability of scoring varied considerably, depending on features in writing that readers could identify. The schemes for reasoning skills were highly confounded by verbal and writing abilities. More objectively defined constructs of written discourse skills must be developed. Four appendices review the scoring schemes, and 13 tables present the findings. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED295995

AERA Presentation

April 1988

Writing Samples Viewed from Different Perspectives: An Approach to Validity

Sybil B. Carlson
Independent Consultant

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

SYBIL B. CARLSON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 011 851

Writing Samples Viewed from Different Perspectives: An Approach to Validity

Sybil B. Carlson

During the present decade we have witnessed the enthusiastic acceptance of writing assessment programs in which writing samples are collected and evaluated. A considerable amount of research has been conducted to address a variety of complex issues. In addition, evaluators have developed a wide range of different scoring schemes designed to obtain the kinds of information that are needed to fulfill the purpose(s) of their writing programs.

Research Objectives

A recent research project provided the opportunity to obtain some preliminary insights about the validity of different scoring schemes that reflected different readers' perceptions when applied to the same set of papers. My concern has been that for too long we have focused our efforts in direct assessment on the reliability of the readers and measures, with only fleeting attention to the validity of the observations being made. High reliability is important to, but does not ensure, validity. Furthermore, it is not sufficient to address validity, the ability to demonstrate that we are measuring what we intend to measure, by claiming that a measure has adequate face and content validity. We also need to demonstrate concurrent validity and construct-related validity.

The research project that will be described serves as an example of the kinds of explorations we need to undertake to investigate the validity of the scores we assign to student papers. Our primary purpose was to determine whether we could objectively describe and identify variables that would meaningfully distinguish reasoning skills couched in written discourse. Other "analytic" scoring schemes have been developed primarily to identify and evaluate writing skills, but with limitations. The validity of many such schemes is not supported adequately. Another objective of the study, funded by the GRE Research Committee, was to gain more information about the writing and reasoning skills being tapped by the verbal and analytical reasoning sections of the Graduate Record Examinations (GRE) General Test. This presentation will emphasize the information gained when we compared scores obtained from different perspectives that were derived from the same writing samples.

Our Research Perspective

Holistic scoring in its various forms recognizes that the singular factors that constitute a piece of writing are elements that work together to make a total impression on the reader.

Holistic scores, though, do not provide sufficient information for diagnostic purposes or for research that will lead us to a better understanding of the interplay among the numerous variables that influence and contribute to a definition of the writing skills being evaluated. As a result, many different scoring schemes have been devised in order to obtain multiple scores. Often these schemes tend to oversimplify and mechanize the evaluation task by adopting a proliferation of labels that reflect a particular reader perspective. These labels do not necessarily reflect the perceptions of another community of readers who do not share the same perspective. As the different scoring schemes continue to be developed, we move further from our attempts to systematically and objectively begin to define the constructs of effective writing in different situational contexts. We lack a common base for demonstrating the validity of our efforts--that we are measuring what we intend to measure.

Methods and Sources of Data

Although my collaborators and I recognized that writing and reasoning skills are confounded in complex ways, we were curious to determine to what extent we could observe reasoning skills as they contributed to written products. Since the papers in the samples we analyzed were written by students entering graduate school, we felt that our approach reflected the realities of this academic setting in which they would be expected to produce written discourse. Academic grades frequently are assigned by educators on the basis of judgments about the quality of student thinking presented in written assignments.

The data base for this study consisted of 406 writing samples prepared by 203 students, primarily nonnative speakers of English (6 Arabic, 73 Chinese, 35 Spanish) and 89 native speakers of English who had recently taken the GRE General Test. The bulk of the data had been collected for a previous writing assessment project, but additional samples were collected to supplement the original sample of native speakers.

Each student prepared two papers on two contrasting topic types, briefly labeled the Space and Farming topics. The Space prompt represents the compare/contrast topic type, whereas the Farming topic represents a chart/graph topic type. These topics were pretested and evaluated in the previous study. (See topics in handout.)

Several experts in writing assessment and linguistics, with experience in research and instruction in reasoning/writing skills and in English as a second language (ESL), developed different scoring schemes with an emphasis on reasoning skills. Four different schemes that required human judgments were applied to the papers as well as the AT&T Writer's Workbench computer analyses. The scores derived from these scoring schemes and

scores on the verbal and analytical reasoning sections of the GRE General Test were related with one another. (See schemes in handout).

The Scoring Schemes

The experts in writing assessment and linguistics were convened for two days to discuss the objectives of the research and to formulate strategies for developing scoring schemes that would emphasize the reasoning skills that might be observed in the writing samples. After working with samples of the papers, each collaborator developed a scheme, appropriately identified by his or her name. The different schemes that were applied to the papers were the following: the Purves/Soter scheme, the Moss scheme, the Reid holistic scoring scheme, the holistic scoring scheme developed in the previous study, and the AT&T Writer's Workbench* computer analyses conducted with the advice of Lawrence Frase.

The schemes developed for this study were intended to focus on the reasoning skills embedded in language as a vehicle for thought, rather than to focus on writing skills per se. The Purves/Soter scheme (developed by Alan Purves and Anna Soter while at the University of Illinois), applied to both the Space and Farming papers, consisted of thirteen separate features subsumed by three major dimensions, content/thinking, organization, and style/tone. Although the labels appear to reflect writing skills, the definitions of the features, with examples, focused on the reasoning skills that might be observed in the papers (Soter, 1985). Readers assigned values on a 1-5 scale, with "1" as the lowest rating, to each of the thirteen features. For reasons aside from this discussion, four readers assigned values to the papers on the Space topic, and two readers, on the Farming topic. The separate subscores assigned by each reader were summed within each dimension, and reader reliabilities were calculated for the three dimensions for each topic. Reader reliabilities (alpha) were acceptably high (in the .80's) for the Space topic, but were lower for the Farming topic (in the .60's). Readers appeared to have greater difficulty in making judgments for this topic. Although the Farming topic reliabilities would not be acceptable in a testing situation, they were adequate for this exploratory analysis, as long as they were taken into account when the data was interpreted. For the final analyses, the scores assigned by each reader, summed within each of the three dimensions for each of the two topics, were averaged to obtain one score per student for each dimension for each topic (six scores per student).

The Moss scheme (developed by Pamela Moss at the University of Pittsburgh) was based on the work of Stephan Toulmin and his colleagues (1984). Related but specific scoring schemes were designed for the two topics since reasoning skills were evidenced

in different ways in response to the specific demands of each topic type. Instead of evaluating the characteristics of the papers, readers indicated the presence of occurrence of each particular feature, resulting in scores that reflected frequencies of occurrence of the major elements of reasoning skills. For the Space topic, four scores were obtained: claims, support, qualifications, and integration. For the Farming topic, six scores were obtained, including four scores that were parallel to the Space topic: graph reading skills, deduction skills, claims, justifications (support), qualifications, and integration. Except for the two unique Farming scores, scores were further derived to eliminate fluency by dividing each score by the total number of ideas offered by each writer.

Estimates of reader reliability (Kuder-Richardson) for the Moss scheme were higher for the Farming topic the for the Space topic, the reverse of the trend for the reliability estimates for the Purves/Soter scheme. For most of these variables, the reliabilities were adequate for this exploration, but not for assessment purposes. For the Farming topic, they were: graph reading, .81; deductions, .62; claims, .71; justifications, .55; qualifications, .45;; and integration, .30. For the Space topic, reader reliabilities were: claims, .59; support, .61; integration, -.02; and qualifications, .51. The extremely low coefficients for integration can be explained by the very small numbers that evidenced this kind of reasoning. This score was dropped from the analyses. For the final analyses, the scores assigned by the two readers were averaged to obtain one score for each element of the Moss scoring scheme on each topic for each writer.

A holistic reasoning scheme was developed by Joy Reid at Colorado State University. A rubric was used by readers to evaluate the papers holistically, with an emphasis on reasoning. Each paper was read by two readers using a six-point scale, with "1" as the lowest rating. The readers reliabilities (coefficient alpha) for both topics were highly acceptable--.88 for the Space topic and .86 for the Farming topic. The scores for the two readers were averaged to yield one holistic reasoning score for each topic for each student.

A major portion of the writing samples were scored holistically with emphasis on writing skills for the previous study. This holistic writing scheme was applied to the additional papers collected, using the original benchmarks, two readers per paper, and a six-point scale in which "1" was the lowest rating. Estimates of reader reliability (coefficient alpha) were acceptably high-- .89 for the Space topic and .92 for the Farming topic. The scores for the two readers were averaged to yield one holistic writing score for each topic for each student.

5/

The Writer's Workbench (WWB) software programs at Colorado State University were used to analyze all writing samples. The output yielded a considerable amount of data describing the semantic and syntactic features of test in numerical forms. Date for a large number of features was obtained, but the relatively unmanageable number were reduced by eliminating highly interdependent, infrequently occurring, and unnecessarily specific features.

The GRE General Test Part Scores

The GRE General Test is composed of three sections: verbal, analytical reasoning, and quantitative. Because the study emphasized verbal reasoning skills, the quantitative scores were not used in the analyses. For research purposes, we calculated part scores for the different subsections of the verbal and analytical reasoning sections of the test. Three parts of the verbal section, sentence completion, discrete verbal, and reading comprehension assess different aspects of language skills. Two parts of the analytical reasoning section represent two important aspects of the reasoning process, analytical reasoning and logical reasoning (Graduate Record Examinations Board, 1986). The analytical reasoning items assess the ability to understand a structure of arbitrary relationships, to deduce new information from the relationships given, and to assess the conditions used to establish the structure of relationships.

In contrast, the task demands of the Space topic require writers to generate a structure of relationships. The Farming topic, on the other hand, offers some structure and vocabulary that writers can use to organize the development of their ideas. The GRE logical reasoning items assess the ability to understand, evaluate, and analyze arguments. Studies validating the analytical reasoning section have indicated that the logical reasoning scores tend to be at least moderately or highly correlated with the verbal section scores, whereas the analytical reasoning scores tend to be correlated moderately with quantitative section scores (Wilson, 1985).

The Data Analyses

From the standpoint of validity, our objective was to determine to what extent the scores obtained from the different instruments supported specific inferences about the appropriateness, meaningfulness, and usefulness of the on the different scoring schemes and scores on the GRE General Test. Separate analyses, as different approaches to viewing the data, were conducted to make the several different comparisons. First, we compared the scores within one scoring scheme to one another to determine the degree of interrelationship or independence that could be obtained by the separate scores. Next, we correlated

all variables in order to observe the interrelationships of the scores among the different scoring schemes and scores on the GRE General Test. Finally, we conducted exploratory factor analyses and stepwise regression analyses to learn more about the information we had gained by using these multiple measures.

The table of correlations among the variables (Table 1) provides most of the information regarding the first two steps of the analyses. The correlations reported are only those that reached a high level of significance. As we interpret the relationships, we looked for the variables that were highly correlated, those that were relatively uncorrelated, and the patterns of correlations among the variables.

Correlations within Schemes

The three Purves/Soter scores were highly correlated within each topic, and somewhat less, though highly correlated, across topics. Thus the three separate scores do not appear to provide independent information because they are so highly related. A factor analysis of the Purves/Soter scores, in fact, indicated that the scores all contributed to one factor.

The scores for the Moss scheme within topics were not highly correlated, and were less highly correlated across topics. The lower correlations across topics were in part influenced by the nature of the scoring schemes, which were designed specifically for each topic. The moderately high correlations among the Moss variables indicate some interdependence, which would be expected, yet the various correlations suggest a relative independence of the separate scores from one another. A factor analysis of the Moss scores indicated that the data consisted of separate factors, and different factors for each of the two topics. This study would require replication, with higher readers reliability estimates, before any interpretations would be reasonable.

Correlations among Schemes

The holistic writing and holistic reasoning scores were highly correlated especially within each topic. These clearly indicate that the two holistic schemes do not provide different information, and that an overall impression score appears to naturally confound writing and reasoning skills. The Purves/Soter scores also had moderately high correlations with the holistic scores, suggesting that the Purves/Soter scores were highly influenced by the writing skills used to convey the reasoning skills of writers.

In contrast, the scores yielded by the Moss scheme demonstrated relatively low to low-moderate relationships with the holistic scores, with the exception of the relatively high

7/

correlations with the support scores for the Space topic and the claims and justifications scores for the Farming topic. These somewhat higher correlations indicate an expected relationships with writing/reasoning skills but also that the scores may be reflecting different skills as well.

The correlations among the Purves/Soter and Moss scores indicate that the Moss scores are relatively independent of the Purves/Soter scores, with the exception of the relatively high correlation of the Moss support score for the Space topic. Because the demands of the Space topic required writers to generate a structure of relationships for which they offered support, these correlations reflect the degree of writing ability that is needed to provide adequate support for an argument in a compare/contrast topic.

Correlations with the GRE General Test Scores

The correlations of the holistic reasoning and holistic writing scores (which essentially assess the same skills) with the GRE verbal scores confirm the relatively high relationship of these scores with verbal abilities. The moderately low correlations of the holistic scores with the analytical reasoning section and relatively high correlations with the logical reasoning section reflects a previously observed phenomenon--that the logical reasoning scores are more highly related to verbal ability than the analytical reasoning scores. This similar pattern of relationships is confirmed by the lower correlations of analytical reasoning with logical reasoning as well.

The scores for the Purves/Soter scheme repeat the same pattern of relationships, although at slightly lower levels, with the verbal, analytical reasoning, and logical reasoning scores of the GRE General Test. These correlations further confirm that the Purves/Soter scheme does not appear to provide sufficient independent information to be used as a separate measure of reasoning skills.

The scores for the Moss scheme reflect even lower correlations with the GRE General Test scores, although some scores (Space support, Farming claims and qualifications) yielded moderate correlations with GRE General Test scores. The Moss scores did not reflect the typical pattern of relationships among the GRE verbal, analytical reasoning and logical reasoning scores, however. These moderate to low correlations indicate, as should be expected, some relationships with verbal and reasoning skills. The sizes of these correlations (Moss with GRE) also suggest that the Moss scheme provides information that is relatively independent of the information obtained from the GRE General Test and other scoring schemes.

8/

Factor Analyses and Regression Analyses

The data for all variables for each topic were factor analyzed (Table 2). The varimax factor analyses should be interpreted with caution, keeping in mind some of the limitations of the data and the fact that broad generalization cannot be drawn on the basis of one study. The different factors obtained for the two topics are not directly comparable, since different variables load on different factors. Generally, however, Factor 1 appears to reflect verbal ability, as measured both directly by the writing samples and indirectly by the GRE General Test. Factor 2 for the Space topic primarily reflects GRE verbal scores, whereas Factor 2 for the Farming topic is dominated by Writer's Workbench (WWB) and Moss variables. Factors 3, 4, and 5 are formed entirely from WWB variables, which appear to be relatively independent of the other variables and suggest that the WWB variables contribute additional information about performance on the writing tasks. In general, Factors 3, 4, and 5 reflect WWB fluency, content, and sentence variety features, respectively.

Stepwise regression analyses were conducted to predict the holistic writing scores and to predict scores for the analytical reasoning and logical reasoning subsections of the GRE analytical reasoning section (Table 3). The regression analyses were purely exploratory, and would need to be replicated to generalize beyond these data; however, they provide interesting information about this set of data. Overall, some relatively predictable features influenced the holistic writing scores. For different language groups on the different, though, slightly different stylistic features contributed to the holistic scores (Table 4). Observations regarding these data are reported in greater detail in the final report for the project (Carlson, 1987, in press), and a recently published chapter (Carlson, 1987).

Implications suggested by the Moss Scheme

Because the Moss variables suggest the potential to obtain additional information about examinee reasoning skills, it is possible to speculate about the kinds of cognitive strategies and concomitant performances that are elicited by the different tasks--the types of essay topics and the GRE General Test items. The two essay topics impose somewhat different demands on the writer, as follows:

- The Farming (chart/graph) topic presents data in tabular form, thus supplying a degree of structure for the relationships among the elements of the data as well as the vocabulary with which ideas can be expressed. A writer can rely extensively on these data to construct statements ("claims") regarding observed

9/

similarities and differences. Further, the writer is asked to summarize by drawing conclusions (final claims) based on the information provided. Although he or she must use cognitive strategies such as analysis and interpretation, the task is relatively concrete. In most papers, in fact, readers observed that the conclusions proposed for the Farming topic represented inferences that did not extend too far beyond the data.

- When responding to the Space (compare/contrast) topic, writers needed to draw from personal experience to develop a structure of logical relationships. The construction of advantages and disadvantages (claims) required independent analysis at a level of abstraction beyond the information given. In addition, writers were asked to render an opinion (a final claim or claims) that also required an evaluation of the similar/different sets of relationships they constructed. Thus, a greater range of cognitive strategies were demanded by this task. In these papers, readers observed that writers appeared to be less fluent and to exhibit more problems with vocabulary and syntax, in contrast to performance on the Farming topic. Sternglass (1986) made similar observations with respect to the different cognitive strategies called for by different types of tasks.

Both essay topics, of course, involve the recall and production of ideas in language, while responding to items on the GRE verbal and analytical reasoning sections requires comprehension of the task, recall of ideas, and the application of strategies for dealing with different kinds of relationships that are assessed through recognition. For the essay production tasks as well as the GRE recognition tasks, assumptions are made regarding the cognitive strategies (e.g., deduction, analysis, evaluation, critical reading) that intervene between the explicit task as it is presented and the resulting performance that is observed. Because the GRE items present different kinds of relationships, however, we can reasonably assume that all examinees have been provided with opportunities to apprehend common sets of relationships. Furthermore, because the GRE item types require an examinee to go beyond the information given as did the Space topic, a correct examinee response serves as an indicator of effective thinking skills.

In the essay tasks, particularly since the instructions did not focus the writers on making their thinking explicit, however, we are not assured that their thinking is either accurately or adequately represented by the ideas produced in writing. The essay task is further confounded by the writer's engagement with

the topic. As Sternglass (1986) notes, a writer's inherent cognitive potential may not be released without the personal commitment to draw upon a full range of cognitive operations leading to more complex thinking and writing.

Among the correlations (Table 3) with the part scores of the GRE verbal section, the support score for the Moss scheme yielded somewhat higher correlations (.35-.40) with the justifications (support) score for the Farming topic. The Space topic, in fact, required evaluation of support provided for the opinion(s) rendered, more so than was required by the Farming topic. Although the correlation was small (.18), the Moss score for graph reading skills was related to the GRE reading comprehension score and to no other GRE verbal or analytical reasoning scores, thus reflecting the demands of the Farming task.

For the Space topic, only the support score was correlated (.21) with the GRE analytical reasoning score; it was more highly correlated (.43) with the logical reasoning score. The support score represents the elaborations, justifications, and implicit and explicit opinions presented by the writer. Similarly, the logical reasoning items on the GRE General Test are intended to assess the ability to understand, analyze, and evaluate arguments, including the recognition of assumptions on which an argument is based and drawing conclusions from premises. Thus, the Moss scheme, when applied to a compare/contrast essay, represents some facets of reasoning that the logical reasoning test is intended to measure--predominantly support for an argument but also the claims and qualifications making up the argument.

For the Farming topic, the Moss scores were reflected equally in correlations among GRE analytical reasoning and logical reasoning scores with deductions, claims, and justifications. The correlations of the Moss qualifications score with the logical reasoning score, however, were moderately high (.42) for the Farming topic and low (.22) for the Space topic, but they were not significant with the analytical reasoning scores. The Farming topic apparently elicited more statements to qualify the claims that could be made about the data, and justifiably so. Since writers were required to create their own structures of relationships for the Space topic, though, their claims were less likely to be stated if they required qualification. Here again, the higher degree of evaluation required by the Space topic is suggested. These correlations of Moss qualifications scores with GRE logical reasoning scores possibly indicate the degree to which the logical reasoning items focus on evaluation through critical analysis or critical reading.

11/

In the varimax factor analysis of all variables (Table 50, the Moss support (justifications), claims, and qualifications scores loaded on one factor with the GRE verbal and both the logical reasoning and analytical reasoning scores for the Farming topic. For the Space topic, the GRE analytical reasoning score contributed to Factor 2, in which the logical reasoning and GRE verbal scores loaded most highly. Again, performance on the compare/contrast Space topic suggests the separate dimensions of reasoning tapped by logical reasoning and analytical reasoning.

Connor (1987) and Connor and Lauer (1987) have used a different adaptation of the Toulmin approach for evaluating the effectiveness of persuasion in compare/contrast essays. Their scheme, in contrast with the more analytical, "frequency count" approach of the Moss scheme, involves ratings of three elements of a paper--claims, data, and warrants. In one analysis, using a holistic (writing) rating as the dependent score, 48 percent of the variance was contributed by the Informal Reasoning score obtained by using their version of the Toulmin model. In another study, the three Informal Reasoning scores (claims, data, warrants) were differentially correlated with the holistic scores for compare/contrast essays; the claim score yielded the highest correlation (.72) with the holistic scores. Since these scores were derived differently than scores for the Moss scheme, direct comparisons cannot be made. However, their research suggests that some form of a Toulmin scheme may provide meaningful information about the reasoning strategies used in written persuasion.

The analyses in our study have demonstrated that the scores derived for the Moss scheme appear to reflect separable components of an argument, particularly for essay topics (compare/contrast) that require the construction of structure and evaluation of support. Because these skills need to be demonstrated in an essay rather than applied when responding to a recognition item, they also may be influenced by the writer's training (e.g., in formal logic), stylistic preferences in responding to the task, and/or degree of engagement with the topic. Such variables that influence performance on a production task would need to be controlled in order to provide examinees with equivalent opportunities to demonstrate their abilities optimally.

The data support the contention that different task demands elicit different kinds of writing performance, and that the writing performance of students representing different first-language groups varies in complex ways in response to these tasks as well. Park (1986), analyzing the data from this study, also confirmed these differential topic effects. He used scores from the Purves/Soter scheme and other measures of textual features, and concluded that significant topic effects were found in elaboration-length measures, the Purves/Soter content/thinking

scores, and the holistic scores. Topic effects were not significant, however, for syntactic complexity measures and the other Purves/Soter scores. Thus, three major results pertinent to the effects of the topic were obtained: (1) writing on the Space writing task evidenced more elaboration (longer essay) and higher quality on the content/thinking score than was evidenced in the Farming task, (2) writing on both tasks exhibited the same degree of syntactic complexity and quality of organization and style/tone (Purves/Soter) scores, and (3) writing on the Farming task yielded higher holistic scores.

Conclusions

As an exploratory investigation that will need to be refined and replicated, its limitations do not warrant undue generalization beyond these data. However, the following conclusions could be drawn: (1) the reliability of scoring schemes applied by human judges can vary considerably, depending on the features in writing the readers are able to identify in samples of actual performance that can vary with the task and native language group; (2) it is difficult to design a scheme for reasoning skills that is not highly confounded by verbal and writing abilities, although one of the schemes showed some potential for deriving independent scores to obtain additional information about writing/reasoning skills; (3) different task demands elicit different kinds of writing performance, which varies in complex ways in response to these tasks; and (4) despite extensive training and general agreement on the features of writing that evidence effective reasoning skills, different communities of readers bring different approaches, labels, and definitions to the observations of these skills.

These observations lead to a recurring, yet basic question in discourse research--do the features observed in written discourse contribute differentially to the quality of writing, depending on the different types and contexts of writing, and are any of these features generalizable across types and contexts? As we pursue future research to investigate this question, we will need to clarify and communicate consistently the expectations of readers, writers, and interpreters of evaluations of written discourse. We need to move toward more objectively defined constructs of written discourse skills that contribute to effective writing in response to different task demands within different contexts.

References

Carlson, S. B. (1987). Cultural differences in writing and reasoning skills. In A. C. Purves (Ed.), Written Communication Annual, Volume 2. Newbury Park, CA: Sage.

Carlson, S. B. (1987, in preparation). Relationships of reasoning and writing skills to GRE analytical ability scores. (GRE Research Report). Princeton, NJ: Educational Testing Service.

Carlson, S. B., Bridgeman, B., Camp, R., & Waanders, J. (1985). Relationships of admission test scores to writing performance of native and nonnative speakers of English. (TOEFL Research Report No. 19, GRE General Report, and ETS Research Report). Princeton, NJ: Educational Testing Service.

Connor, U. (1987). Personal communication.

Connor, U., & Lauer, J. (1987, in press). Crosscultural variation in persuasive student writing. In A. C. Purves (Ed.), Written communications annual, Vol. 2. Newbury Park, CA: Sage.

GRE guide to the use of the Graduate Record Examinations program, 1986-87 (1986). Princeton, NJ: Educational Testing Service.

Park, Y. M. (1986). The influence of task upon writing performance. Unpublished doctoral dissertation: University of Illinois at Urbana-Champaign.

Soter, A. (1985). GRE/TOEFL scoring criteria. Unpublished manuscript. Urbana, IL: Curriculum Laboratory, University of Illinois.

Sternglass, M. (1986). Commitment to writing and complexity of thinking. Journal of Basic Writing, 5 (1), 77-88.

Toulmin, S., Rieke, R., & Janik, A. (1984). An introduction to reasoning, second edition. New York, NY: Macmillan.

Wilson, K. (1985). The relationship of GRE General Test item-type part scores to undergraduate grades. (GRE Professional Report, GRE No. 81-22P). Princeton, NJ: Educational Testing Service.

AERA Presentation

April 1988

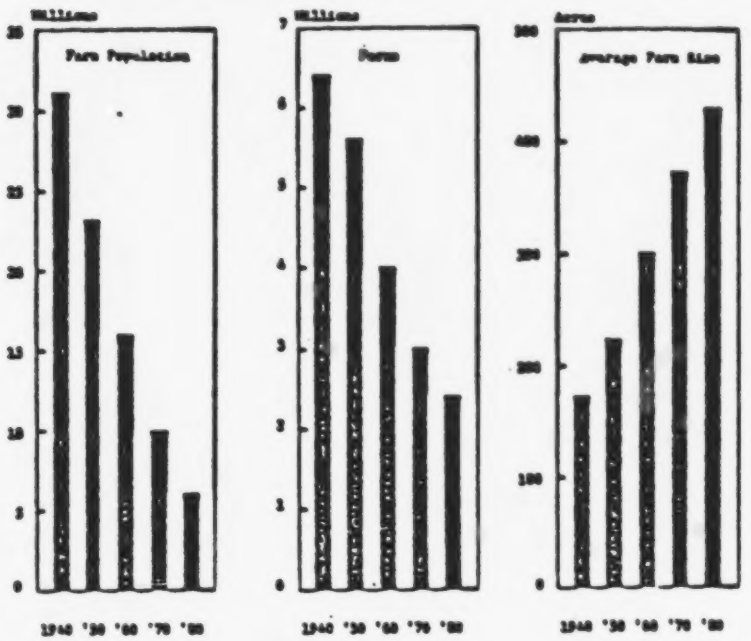
Writing Samples Viewed from Different Perspectives: An Approach to Validity

Appendices and Tables

Sybil B. Carlson
Independent Consultant

TIME - 30 MINUTES

CHANGES IN FARMING IN THE U.S.: 1940 - 1980



Suppose that you are writing a report in which you must interpret the three graphs shown above. Write the section of that report in which you discuss how the graphs are related to each other and explain the conclusions you have reached from the information in the graphs. Be sure the graphs support your conclusions.

THIS SPACE MAY BE USED FOR NOTES.

TIME - 30 MINUTES

Some people say that exploration of outer space has many advantages; other people feel that it is a waste of money and other resources. Write a brief essay in which you discuss each of these positions. Give one or two advantages and disadvantages of space exploration, and explain which position you support.

THIS SPACE MAY BE USED FOR NOTES.

Appendix B

Purves/Soter Scoring Scheme

A. Content/Thinking

This dimension concerns what is written and the way it reflects the writer's manipulation of the subject. It is divided into a number of specific aspects which should be scored separately.

1. Adequacy of information presented refers to the degree to which all of the relevant information from the stimulus is contained in the text.

2. Richness of additional information refers to the use of additional information to that which is in the stimulus (e.g., information drawn from a variety of sources such as reading or general knowledge) and may be seen as the amount of relevant allusion.

3. Relationships drawn refers to the degree to which the text shows that connections have been made between or among the various items of information and the validity and/or complexity of the relationships.

4. Inferences made refers to the number and depth of interpretations (causal, resultant, comparative, contrastive, extrapolative) that the writer makes beyond the information in the stimulus and or from the outside.

5. Synthesis refers to the degree to which the writer appears to draw together the information, relationships, and inferences into a single or complex generalization.

6. Evaluation refers to the degree to which the writer appears to make judgements as to the relative merit of particular relationships, inferences, or syntheses and the degree to which applicable criteria are used.

7. Consideration of alternatives refers to the extent to which the writer appears to admit the possibility of alternative or counter arguments (as in Space) or interpretations (as in Farms) and either accepts them as admissable or rebuts them.

B. Organization

This dimension concerns the structure of the written text both as a whole text and in its various parts.

8. Framing refers to the degree to which the writer presents a context for the content in such a manner that there is an apparent beginning, middle, and end. There need not be a formal introduction or conclusion but there should be a clear sense that the composition begins and ends appropriately.

9. Grouping refers to the degree to which the writer joins (in paragraphs or through some other means) the various pieces of information, relationships or inferences). Inadequate combining would be represented by a list in which there is no discernible pattern or system of combining bits of information.

10. Unity refers to the degree to which the writer appears to have both grouped appropriate information, relationships, inferences and excluded extraneous matter or thoughts. The reader may see this in units such as paragraphs and/or in the composition as a whole.

C. Style/Tone

This dimension refers to the manner in which the composition is presented, and particularly to the degree to which that manner matches the conventions of academic discourse in the United States.

11. Objectivity refers to the use of impersonal and detached language as opposed to personal and emotional language. An objective composition may use the first person singular occasionally, but usually as a qualifier.

12. Tentativeness refers to the use of semantic hedges and qualifiers that are often considered appropriate to academic writing. In the responses to the particular stimuli of these assignments, the better composition will probably not use extreme or dogmatic language.

13. Metalanguage refers to the use of markers (e.g., "however") to indicate the relationship between propositions and paragraphs. These markers would include connectives and markers which illustrate such relationships as cause, effect, comparison, contrast classification, definition or hypothesis. The markers would signal the sequence and structure of the composition.

Appendix C

Moss Scoring Scheme

Space Topic

ID _____ Reader _____

(+ = existence; blank = non-existence)

Advantages

___ Brief descriptor_____

___ clear; non-ambiguous

___ accurate fact or reasonable inference; relevant

___ elaborated with examples (or a more specific extension of the advantage without justification/explanation)

___ justified/explained (e.g., "this is an advantage because...")

___ qualified/rebutted (i.e., shown not to be an advantage)

___ builds on previous advantage (to justify/support current advantage)

___ integrates as evidence at least two previous advantages

[continues for all descriptors]

[on subsequent pages]

Disadvantages

___ Brief descriptor_____

___ clear; non-ambiguous

___ accurate fact or reasonable inference; relevant

___ elaborated with examples (or a more specific extension of the disadvantage without justification/explanation)

___ justified/explained (e.g., "this is a disadvantage because...")

___ qualified/rebutted (i.e., shown not to be a disadvantage)

___ builds on previous disadvantage (to justify/support current disadvantage)

___ integrates as evidence at least two previous disadvantages

---

Opinion/Point of View

___ Brief descriptor_____

___ explicit summary statement of position made

___ position clear, non-ambiguous (scoring can be based on an explicit statement of position or on a position implied in presentation of advantages and disadvantages)

___ criteria explicit (advantages/disadvantages explicitly connected to opinion)

___ criteria implicit (advantages/disadvantages implicitly connected to opinion)

___ advantages/disadvantages rebutted or qualified in ways consistent with opinion (explicit recognition of competing priorities)

C/3

Abridged Version of Moss Scoring Scheme

Farming Topic

ID _____ Reader_____

(+ = existence; blank = non-existence)

Descriptions of/Deductions from Graphs

___ accurate description of graph 1 (Farm population decreased)

___ specific description of graph 1 (at least three pieces of information on two data points)

___ accurate description of graph 2 (Number of farms decreased)

___ specific description of graph 2 (at least three pieces of information on two data points)

___ accurate description of graph 3 (Size of farms increased)

___ specific description of graph 3 (at least three pieces of information on two data points)

___ relationship between graphs 1 and 2 explicit

___ new variable or quantity deduced from graphs 1 and 2

___ relationship between graphs 1 and 3 explicit

___ new variable or quantity deduced from graphs 1 and 3

___ relationship between graphs 2 and 3 explicit

___ new variable or quantity deduced from graphs 2 and 3

___ relationship between graphs 1, 2, and 3 explicit

___ new variable or quantity deduced from graphs 1, 2, and 3

Conclusion(s)

Conclusions include statements that draw inferences from the graphs (e.g., of cause, consequence, prediction) which may require additional information or justification for support.

___ Brief descriptor _____

_____

___ clear; non-ambiguous

___ reasonable inference

___ evidence provided or requested/justification given (in addition to data from the graphs)

___ elaborated with examples (or a more specific conclusion/extension of the conclusion without justification)

___ obvious connection made between data from graphs and conclusion

___ qualified/possible rebuttal mentioned (e.g., labeled as "likely" or "possible" conclusion, rather than "the" conclusion; or as a likely conclusion "unless" certain conditions change)

___ builds on previous conclusion(s) (to justify/support current conclusion)

___ integrates as evidence at least two previous conclusions

[continues for however many conclusions are presented]

Appendix D
Reid Holistic Reasoning Schemes

Space Topic

General Considerations

Content

Quality of analysis
of detail
of logic
of communication of ideas
of focus

Length: breadth, depth, and appropriateness of detail; limits and focuses topic

Process/Organization

May not fulfill expectations of the U.S. academic audience, but must have the overall impact of completeness, and the sequencing of ideas in a logical manner

Describes/States the problem: introduction (may be implied)
Provides statement of focus (topic sentence): choosing an alternative
Develops ideas by specific detail, examples, explanation
Observes/Explains the choices
Analyzes the value/benefits; provides supporting analysis with detailed material
Concludes with summary, solution, question, prediction; based on earlier arguments and/or examples

Space

Scoring

Pluses that contribute to the holistic evaluation

Unusually tight (vs. unusually loose)
Unusually rich (vs. paucity of development or examples); may develop the argument with examples
Originality within parameters of task (vs. regurgitating topic)

High Range Papers (5-6 on a 1-6 scale)

Weighted persuasive presentation (or perhaps equally presented both sides)
Adequate, appropriate support (facts, examples, physical description, and/or personal experience)
High degree of overall fluency (?)
Must include one or more of the +'s

Note: some high papers don't take a stand on the alternatives; sometimes they present a balanced explanation of both sides, then leave the choice to the reader (they don't presume to choose)

Middle Range Papers (4-5 on a 1-6 scale)

May repeat the assignment as the statement of introduction
Has a sense of direction and some detail, but focus or support is thin or non-existent
May state opinions without support
Overall fluency flawed (?)
For a 4: basic elements are there without +'s
For a 3: some basic elements are there without +'s

Low Range Papers (1-2 on a 1-6 scale)

Organization and focus flawed seriously or non-existent
Little or no knowledge of what comparison/contrast demands
Little or no ability to assume and produce the task
Overall fluency extremely limited or non-existent (?)
Serious lack of basic elements

D/3

Holistic Reasoning Scoring Scheme

Farming Topic

General Considerations

Content

Quality of analysis: detail, logic, communication of ideas, focus

Length: breadth, depth, and appropriateness of detail; limits and focuses topic

Process/Organization: May not fulfill expectations of the U.S. academic audience, but must have the overall impact of completeness, and the sequencing of ideas in a logical manner

- States the problem or issue (may be implied)
- Describes the chart(s): overall description
- Discusses the chart(s): more specific details of description that links the charts
- Analyses the chart(s): integrates the information, provides causal connections
- Concludes: possible solution, summary, provocative questions, predictions; based on earlier or following arguments and/or examples

Pluses (+) that contribute to the holistic evaluation:

- Unusually tight (vs. unusually loose)
- Unusually rich (vs. paucity of development or examples); may develop the argument with examples
- Originality within parameters of task (vs. restating topic)

High Range Papers (5-6 on a 1-6 scale)

- Clear organization; may make direct reference to the chart(s) and/or assignment
- Careful analysis; adequate support for analysis
- Overall fluency high —focus on presentation of ideas vs. vocabulary
- Must include one of more of the +'s
  - Note: some high papers don't analyze (don't presume to do so with so little information); some high papers don't describe in great detail the charts—they get right to the business of analysis (implying a degree of audience background)

Middle Range Papers (4-5 on a 1-6 scale)

- Clear, detailed description; perhaps limited discussion
- Some analysis, but little or no support for the analysis
- Overall fluency flawed
- A score of 4: basic elements are there without +'s
- A score of 3: some basic elements are there without +'s
  - Note: at least 2 of these elements are usually present in middle-range papers, often more

Low Range Papers (1-2 on a 1-6 scale)

- Description but little or no discussion or analysis
- Organization is not evident; simple chronology at best
- Little or no detail; Serious lack of elements
- Summary follows immediately, if at all
- Overall fluency is very limited or non-existent
- Serious mechanical errors interfere with basic communication of ideas

Table 1

Correlations of Holistic Scores Total* and GRE Item Type Scores

(sample of 132 cases)

Scores	Hol.	GRE Scores: Verbal SC	Verbal DV	Verbal RC	Quantitative QC	Quantitative M	Quantitative DI	Analytical AR	Analytical LR
GRE Verbal									
Sentence Completion (SC)	.68								
Discrete Verbal (DV)	.67	.64							
Reading Comprehension (RC)	.70	.70	.64						
GRE Quantitative									
Quantitative Comparisons (QC)	-.22	-.26	-.30	-.12					
Discrete Math (M)	-.31	-.28	-.36	-.26	.76				
Data Interpretation (DI)	-.09	-.03	-.08	.00	.64	.59			
GRE Analytical									
Analytical Reasoning (AR)	.23	.15	.17	.24	.46	.35	.50		
Logical Reasoning (LR)	.64	.65	.50	.67	-.09	-.18	.02	.24	

*Holistic scores averaged over four writing samples

Table 2

Means and Standard Deviations of Variables in Analysis
(N=203)

Variables	Space Topic (29 variables) Mean	S.D.	Farming Topic (30 variables) Mean	S.D.
Holistic Writing	4.23	1.57	4.34	1.58
Holistic Reasoning	3.49	1.21	3.53	1.33
Purves/Soter Scheme				
Content	18.44	4.80	17.72	4.94
Organization	8.23	2.14	8.01	2.50
Style and Tone	7.56	2.12	7.70	1.99
Moss Scheme-Space				
Claims	5.88	2.20	—	—
Support	2.67	1.53	—	—
Qualifications	.45	.54	—	—
Moss Scheme--Farming				
Graph Reading	—	—	3.70	1.36
Deductions	—	—	.89	1.00
Claims	—	—	2.70	1.21
Justifications	—	—	1.63	1.13
Qualifications	—	—	.15	.25
Writer's Workbench Variables				
Number of suggestions	6.14	4.58	6.35	5.06
No. spelling errors	2.81	2.55	2.09	2.47
% Vague words	6.71	2.54	5.22	2.67
Number to check	2.50	1.77	2.07	1.66
Av. sentence length	21.10	6.69	22.11	7.50
% shorter sentences	27.40	14.49	28.12	13.34
% longer sentences	11.58	8.40	11.42	9.25
% to be verbs	81.14	11.86	75.21	18.74
% passive voice	11.32	9.09	11.56	10.46
% nominalizations	3.56	1.42	2.80	1.32
Av. word length	4.70	.31	4.65	.27
% content words	54.88	3.80	57.48	4.00
Av. length content words	6.20	.46	5.86	.67
% prepositions	11.00	2.42	13.33	3.35
% adjectives	15.07	3.14	17.60	3.60
% abstract	3.55	1.44	2.90	1.32

GRE General Test Scores	Mean	S.D.
Verbal		
Sentence completion	6.99	3.85
Discrete verbal	19.02	8.89
Reading comprehension	11.95	5.60
Analytical Reasoning		
Analytical reasoning	21.25	6.61
Logical reasoning	5.85	2.77

Table 3

Correlations of Scoring Schemes and GRE Part Scores (N=203)
(.001 level of significance)

			GRE General Test					Purves/Soter Scheme						Moss Scheme								Holistic Schemes			
								Space			Farming			Space			Farming					HR		HV	
			SC	DV	RC	AR	LR	CT	O	ST	CT	O	ST	C	S	Q	G	D	C	J	Q	S	F	S	F
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
GRE V	1	Sent Comp																							
	2	Disc V	.81																						
	3	Read Comp	.80	.77																					
GRE AR	4	Analyt R	.39	.39	.43																				
	5	Logical R	.76	.68	.77	.48																			
Purves/	6	Cont/Think	.67	.66	.72	.33	.62																		
Soter	7	Organization	.69	.66	.71	.29	.62	.93																	
Space	8	Style & Tone	.70	.69	.72	.33	.65	.91	.90																
Purves/	9	Cont/Think	.64	.61	.68	.33	.59	.73	.74	.73															
Soter	10	Organization	.62	.59	.66	.34	.61	.72	.73	.72	.91														
Farming	11	Style & Tone	.65	.64	.69	.34	.64	.76	.76	.77	.88	.88													
Moss	12	Claims	.31	.35	.35		.28	.52	.52	.52	.37	.33	.36												
Space	13	Support	.55	.57	.58	.21	.43	.72	.71	.71	.58	.56	.57	.55											
	14	Qualification	.32	.38	.35		.22	.32	.33	.35	.31	.30	.30	.17*	.32										
Moss	15	Graph			.18*			.26	.25	.24	.22	.31	.28		.26										
Farming	16	Deduction	.25	.18*	.28	.22	.26	.24	.24	.21	.17*	.24	.26			.23	.19								
	17	Claims	.40	.46	.42	.35	.37	.41	.40	.41	.51	.45	.39	.22	.30										
	18	Justification	.35	.36	.40	.20	.30	.37	.38	.39	.57	.51	.41	.22	.31				.64						
	19	Qualification	.44	.41	.38		.42	.42	.42	.39	.38	.41	.48	.18*	.31	.28	.17*		.19*						
Holistic	20	Space	.68	.72	.70	.37	.60	.81	.81	.79	.69	.70	.70	.63	.77	.42	.24	.24	.43	.39	.41				
Reasoning	21	Farming	.72	.69	.74	.41	.64	.74	.75	.75	.81	.80	.78	.35	.59	.34	.33	.25	.64	.61	.47	.74			
Holistic	22	Space	.76	.74	.76	.35	.70	.82	.81	.80	.70	.71	.74	.44	.65	.34	.23	.27	.37	.34	.43	.80	.74		
Writing	23	Farming	.72	.71	.76	.37	.69	.78	.76	.78	.78	.77	.79	.45	.60	.34	.30	.26	.48	.46	.39	.76	.82	.85	

*at .01 level.

Table 4

Varimax Factor Analysis of Writer's Workbench Variables
(N = 203)

Farming Topic		Space Topic	
Factor 1	Loading	Factor 1	Loading
% content words	.99	% content words	.99
% number of adjectives	.65	% adjectives	.69
Number of spelling errors	-.22		
Average word length	.40		
Factor 2		Factor 2	
% shorter sentences	.79	% shorter sentences	.88
Average sentence length	.60	Average sentence length	.61
% longer sentences	.52	% longer sentences	.53
Number to check	.29	Number of spelling errors	.32
% prepositions	.25	Average word length	.55
Factor 3		Factor 3	
Average word length	.76	Avg. length content words	.89
Average length content words	.75	Average word length	.79
% nominalizations	.37	% prepositions	.62
% vague words	-.35	% vague words	-.59
% prepositions	.26	% passive voice	.56
		% to be verbs	.47
		Number to check	.35
		Average sentence length	.34
		% nominalizations	.31
Factor 4		Factor 4	
% passive voice	.71	% nominalizations	-.62
% to be verbs	.37	Number of suggestions	.44
Number of suggestions	.30	% abstract	-.15
% abstract	-.14	Number to check	.35
(accounting for 35.5% variance)		(accounting for 47.9% variance)	

Table 5

Varimax Factor Analysis of All Variables
(N = 203)

Farming Topic		Space Topic	
Factor 1	Loading	Factor 1	Loading
Holistic reasoning	.92	Purves/Soter content	.86
Purves/Soter content	.89	Purves/Soter organization	.84
Holistic writing	.87	Holistic reasoning	.82
Purves/Soter organization	.86	Purves/Soter style & tone	.82
Purves/Soter style & tone	.85	Moss support	.78
GRE reading comprehension	.82	Holistic writing	.68
GRE sentence completion	.79	Moss claims	.63
GRE discrete verbal	.77	WWB number to check	.48
GRE logical reasoning	.74	WWB number of suggestions	.41
Moss justification	.63	Moss qualifications	.31
Moss claims	.62	GRE logical reasoning	.42
WWB number of suggestions	.54	GRE sentence completion	.50
GRE analytical reasoning	.44	GRE reading comprehension	.54
WWB number to check	.44	GRE discrete verbal	.51
Moss qualifications	.43		
WWB % to be verbs	.22		
WWB % abstract	-.15		
WWB average word length	.36		
Factor 2		Factor 2	
WWB % prepositions	.58	GRE logical reasoning	.71
WWB % vague words	-.56	GRE sentence completion	.70
Moss deductions	.40	GRE reading comprehension	.63
Moss graph reading	.39	GRE discrete verbal	.62
Moss claims	-.45	GRE analytical reasoning	.44
Moss justifications	-.33	Holistic reasoning	.30
		Purves/Soter style & tone	.34
		Holistic writing	.48
Factor 3		Factor 3	
WWB average word length	.84	WWB av. length content words	.83
WWB av. length content words	.71	WWB average word length	.72
WWB % nominalizations	.34	WWB % prepositions	.50
		WWB % vague words	-.49
		WWB % nominalizations	.47
		WWB % passives	.46
		WWB % to be verbs	.42
Factor 4		Factor 4	
WWB % content words	.95	WWB % content words	.88
WWB % adjectives	.64	WWB % adjectives	.71
WWB % passive voice	-.18	WWB average word length	.52
		WWB % nominalizations	.34
Factor 5		Factor 5	
WWB % shorter sentences	.78	WWB % shorter sentences	.88
WWB average sentence length	.56	WWB average sentence length	.61
WWB % of longer sentences	.54	WWB % longer sentences	.59
WWB number of spelling errors	.26	WWB number of spelling errors	.28
		WWB % abstract	-.08
(accounting for 48.4% variance)		(accounting for 56.0% variance)	

Table 6

Means and Standard Deviations of Variables in Analysis for the Three Language Groups

Space Topic

Variables	Chinese (N= 73) Mean	SD	English (N= 89) Mean	SD	Spanish (N= 35) Mean	SD
Holistic Writing	2.78	1.05	5.60	.69	3.70	1.10
Holistic Reasoning	2.48	.79	4.47	.86	3.03	.61
Purves/Soter Scheme						
Content	14.18	3.12	22.30	3.00	17.53	3.27
Organization	6.40	1.55	9.92	1.28	7.65	1.53
Style and Tone	5.71	1.08	9.37	1.40	6.69	1.32
Moss Scheme						
Claims	4.79	1.93	6.88	2.23	5.52	1.46
Support	1.60	1.08	3.69	1.33	2.23	.97
Qualifications	.24	.34	.70	.64	.22	.32
Writer's Workbench Variables						
Number of suggestions	4.04	3.26	8.21	5.08	5.09	3.18
No. spelling errors	2.97	2.17	2.47	2.74	3.34	2.83
% Vague words	8.24	2.92	5.59	1.83	6.56	1.69
Number to check	1.50	1.29	3.38	1.80	2.34	1.49
Av. sentence length	18.66	7.61	21.90	4.81	23.96	7.27
% shorter sentences	23.28	16.37	28.13	11.40	34.23	15.09
% longer sentences	9.29	9.19	12.29	6.15	13.51	9.84
% to be verbs	77.18	12.33	84.22	10.40	80.97	11.59
% passive voice	6.11	6.81	15.26	8.76	11.83	8.65
% nominalizations	3.68	1.63	3.52	1.21	3.51	1.48
Av. word length	4.49	2.71	4.90	2.26	4.61	2.73
% content words	54.60	3.56	56.11	3.02	52.65	4.47
Av. length content words	5.88	4.02	6.48	3.23	6.18	4.39
% prepositions	9.26	2.28	11.99	1.59	11.72	2.60
% adjectives	14.53	2.85	16.20	2.84	13.54	3.47
% abstract	3.55	1.67	3.50	1.17	3.59	1.44

(GRE General Test Scores on Table 7)

Table 7

Means and Standard Deviations of Variables in Analysis for the Three Language Groups

Farming Topic

Variables	Chinese (N= 73) Mean	SD	English (N= 89) Mean	SD	Spanish (N= 35) Mean	SD
Holistic Writing	2.92	1.04	5.78	.52	3.61	1.10
Holistic Reasoning	2.49	.78	4.66	.84	2.81	.89
Purves/Soter Scheme						
Content	13.58	3.00	21.67	3.50	16.29	3.14
Organization	6.08	1.49	10.00	1.79	7.00	1.90
Style and Tone	5.97	1.16	9.35	1.28	6.99	1.34
Moss Scheme						
Graph Reading	3.51	1.30	4.02	1.16	3.30	1.77
Deductions	.61	.72	1.22	1.17	.64	.82
Claims	2.20	1.15	3.23	1.02	2.46	1.25
Justifications	1.14	.88	2.14	1.16	1.34	.97
Qualifications	.03	.10	.26	.30	.11	.18
Writer's Workbench Variables						
Number of suggestions	3.49	3.29	9.00	5.31	5.60	3.63
No. spelling errors	1.79	1.79	1.82	2.46	3.49	3.25
% Vague words	5.59	3.27	4.82	2.03	5.76	2.67
Number to check	1.18	1.21	2.71	1.57	2.37	1.90
Av. sentence length	19.68	6.89	22.33	6.53	24.93	6.24
% shorter sentences	26.10	14.68	27.91	10.86	32.23	14.04
% longer sentences	10.22	9.11	11.21	8.12	14.83	10.87
% to be verbs	71.83	19.91	79.64	16.38	69.54	20.17
% passive voice	10.10	12.36	13.74	9.05	8.00	7.06
% nominalizations	2.96	1.30	2.53	1.25	3.17	1.44
Av. word length	4.55	2.60	4.78	2.23	4.51	2.50
% content words	57.99	3.87	58.34	3.53	54.37	3.50
Av. length content words	5.63	7.38	6.06	6.58	5.82	3.37
% prepositions	12.22	3.77	14.13	2.47	13.77	3.61
% adjectives	17.17	4.07	18.46	3.05	16.39	3.49
% abstract	3.15	1.40	2.67	1.22	2.93	1.39
GRE General Test Scores						
Verbal						
Sentence completion	3.60	1.63	10.03	2.94	6.29	2.81
Discrete verbal	11.18	4.87	25.87	6.64	18.29	5.69
Reading comprehension	6.62	3.03	16.63	3.33	10.86	4.01
Analytical Reasoning						
Analytical reasoning	19.89	6.09	24.02	6.15	17.26	5.99
Logical reasoning	3.74	1.47	7.93	2.26	4.86	2.17

Table 8

Stepwise Regression Analyses Predicting Holistic Writing Scores from 21 Writer's Workbench Variables (before final reduction to 16 variables)

Space Topic

Total Sample (N=203)

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	Number of words	.55	.54	.74
2	Average word length	.66	.66	.37
3	Number of spelling errors	.69	.69	-.17
4	Percent passives	.71	.70	.16
5	Percent nominalizations	.71	.71	-.10
6	Kincaid readability	.72	.71	-.08
7	Percent simple sentences	.73	.72	-.10
8	Percent vagueness	.73	.72	-.09
9	Percent adjectives	.74	.73	.08

(F= 60.22 with 9, 193 df, p=.001)

Chinese Sample (N=73)

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	Number of words	.19	.18	.44
2	Percent abstract	.26	.24	-.25
3	Percent longer sentences	.32	.29	-.25
4	Percent vague words	.36	.32	-.21
5	Percent adjectives	.39	.35	.17
6	Average sentence length	.43	.37	-.22

(F= 8.17 with 9, 193 df, p=.001)

English Sample (N=89)

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	Number of words	.09	.08	.30
2	Percent adjectives	.14	.12	.22*

(F= 6.79 with 9, 193 df, p=.001)

Spanish Sample (N=35)

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	Percent passives	.28	.26	.53
2	Percent shorter sentences	.51	.48	.48

(F= 16.70 with 9, 193 df, p=.001)

*Asterisks indicate variables with negative beta weights that were positively correlated with the criterion

Table 9

Stepwise Regression Analyses Predicting Holistic Writing Scores from 21 Writer's Workbench Variables (before final reduction to 16 variables)

Farming Topic

Total Sample (N=203)

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	Number of words	.45	.45	.67
2	Average word length	.55	.54	.31
3	Percent prepositions	.58	.58	.19
4	Percent nominalizations	.61	.60	-.17
5	Number of suggestions	.62	.61	.13
6	Number of sentences	.64	.63	.23
7	Percent vague words	.64	.63	-.10
8	Av. length content words	.65	.64	.10
9	Number of spelling errors	.66	.64	-.09

(F= 40.94 with 9, 193 df, p=.001)

Chinese Sample (N=73)

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	Number of sentences	.11	.10	.34
2	Percent prepositions	.28	.26	.43
3	Av. length content words	.32	.29	.21
4	Number of spelling errors	.35	.31	-.19

(F= 9.27 with 9, 193 df, p=.001)

English Sample (N=89)

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	Number of words	.22	.22	.47
2	Percent shorter sentences	.31	.29	.29
3	Percent content words	.35	.33	.21
4	Percent to be verbs	.37	.34	.15

(F= 12.22 with 9, 193 df, p=.001)

Spanish Sample (N=35)

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	Number of sentences	.45	.43	.67
2	Number of suggestions	.51	.48	.25

(F= 16.42 with 9, 193 df, p=.001)

Table 10

Stepwise Regression Analyses Predicting GRE Analytical Reasoning Scores from All Variables (GRE General Verbal Subscores and Writing Sample Scoring Schemes) for Total Sample (N=203)

Space Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE reading comprehension	.18	.18	.43
2	WWB percent passives	.21	.20	-.17*
3	WWB percent shorter sentences	.23	.22	-.14
4	Holistic reasoning	.24	.23	.18

(F= 15.90 with 4,198 df, p=.001)

Farming Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE reading comprehension	.18	.18	.43
2	WWB average sentence length	.22	.21	-.20
3	Moss claims	.24	.23	.17
4	Moss deductions	.27	.26	.18
5	WWB av. length content words	.29	.27	-.13

(F= 15.98 with 4,198 df, p=.001)

Stepwise Regression Analyses Predicting GRE Logical Reasoning Scores from All Variables (GRE General Verbal Subscores and Writing Sample Scoring Schemes) for Total Sample (N=203).

Space Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE reading comprehension	.60	.60	.77
2	GRE sentence completion	.65	.65	.40
3	Holistic writing	.66	.66	.13
4	Moss support	.67	.66	-.13*
5	Purves/Soter style and tone	.68	.67	.14
6	Moss qualifications	.68	.67	-.07*

(F= 69.43 with 4,198 df, p=.001)

Farming Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE reading comprehension	.60	.60	.77
2	GRE sentence completion	.65	.65	.40
3	Holistic writing	.66	.66	.15
4	WWB percent to be verbs	.67	.66	.07

(F= 99.98 with 4,198 df, p=.001)

*Asterisks indicate variables with negative beta weights that were positively correlated with the criterion

Table 11

Stepwise Regression Analyses Predicting GRE Analytical Reasoning and Logical Reasoning Scores from All Variables (GRE General Verbal Subscores and Writing Sample Scoring Schemes) by Language Groups

Chinese Sample (N=73)

Analytical Reasoning as Dependent Variable

Space Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	WWB percent passive voice	.05	.04	-.23
2	Holistic reasoning	.09	.06	.20
3	Moss support	.14	.10	-.32
4	WWB percent short sentences	.17	.12	-.19

(F= 3.50 with 4, 67 df, p=.001)

Farming Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	Moss claims	.12	.11	.35
2	WWB percent short sentences	.18	.15	-.24
3	WWB percent adjectives	.24	.20	-.24
4	WWB percent passive voice	.29	.24	-.23
5	Moss graph reading skills	.33	.28	.21

(F= 6.50 with 5, 66 df, p=.001)

Logical Reasoning as Dependent Variable

Space Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE sentence completion	.16	.15	.41
2	WWB average word length	.22	.20	-.24
3	GRE reading comprehension	.29	.26	.27
4	Holistic writing	.33	.29	.21
5	WWB percent abstract	.36	.32	.20
6	Moss support	.40	.34	-.24*
7	WWB percent longer sentences	.43	.37	.18

(F= 6.84 with 7, 64 df, p=.001)

Farming Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE sentence completion	.16	.15	.41
2	GRE reading comprehension	.21	.19	.23
3	WWB average sentence length	.25	.22	-.21
4	Purves/Soter content/thinking	.29	.25	-.21
5	WWB average word length	.34	.29	-.22
6	Moss graph reading skills	.37	.31	.18
7	WWB percent abstract	.40	.33	.17

(F= 5.99 with 7, 64 df, p=.001)

*Asterisks indicate variables with negative beta weights that were positively correlated with the criterion

Table 12

Stepwise Regression Analyses Predicting GRE Analytical Reasoning and Logical Reasoning Scores from All Variables (GRE General Verbal Subscores and Writing Sample Scoring Schemes) by Language Groups

English Sample (N= 89)

Analytical Reasoning as Dependent Variable

Space Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE sentence completion	.14	.13	.38
2	Moss qualifications	.20	.18	-.24
3	WWB percent longer sentences	.25	.23	-.23
4	WWB percent nominalizations	.30	.27	-.23

(F= 9.38 with 4, 84 df, p=.001)

Farming Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE sentence completion	.14	.13	.38
2	WWB av. length content words	.24	.22	-.30
3	WWB number to check	.27	.25	.19

(F= 10.99 with 3, 85 df, p=.001)

Logical Reasoning as Dependent Variable

Space Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE sentence completion	.33	.33	.58
2	GRE reading comprehension	.39	.38	.33
3	Moss qualifications	.42	.40	-.17

(F= 21.41 with 3, 85 df, p=.001)

Farming Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE sentence completion	.33	.33	.58
2	GRE reading comprehension	.39	.38	.33
3	Moss graph reading	.42	.40	-.14
4	Moss qualifications	.44	.41	.16
5	WWB percent content words	.46	.43	.17
6	Moss justifications	.48	.44	-.14

(F= 13.15 with 6, 82 df, p=.001)

Table 13

Stepwise Regression Analyses Predicting GRE Analytical Reasoning and Logical Reasoning Scores from All Variables (GRE General Verbal Subscores and Writing Sample Scoring Schemes) by Language Groups

Spanish Sample (N=35)

Analytical Reasoning as Dependent Variable

Space Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE reading comprehension	.44	.42	.66
2	WWB percent nominalizations	.53	.50	-.30
3	WWB percent adjectives	.61	.57	.28

(F= 16.00 with 3, 31 df, p=.001)

Farming Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE reading comprehension	.44	.42	.66
2	Moss deductions	.55	.52	.33

(F= 19.29 with 2, 32 df, p=.001)

Logical Reasoning as Dependent Variable

Space Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	reading comprehension	.35	.33	.59
2	Holistic writing	.44	.40	.32
3	WWB percent vague words	.56	.52	.37
4	Moss support	.61	.56	-.23
5	Purves/Soter style and tone	.69	.64	.31
6	WWB percent nominalizations	.73	.67	.19
7	WWB av. length content words	.77	.71	-.28

(F= 12.76 with 7, 27 df, p=.001)

Farming Topic

Step	Writer's Workbench Independent Variables	R Squared	Adjusted R Squared	beta
1	GRE reading comprehension	.35	.33	.59
2	Holistic writing	.49	.46	.40
3	WWB percent content words	.55	.50	-.25
4	Holistic reasoning	.59	.54	-.26*
5	Purves/Soter style and tone	.66	.60	.30
6	GRE sentence completion	.71	.65	.29
7	WWB average sentence length	.75	.69	.22
8	WWB percent abstract	.79	.72	.22
9	Moss justifications	.81	.74	.22

(F= 11.93 with 9, 25 df, p=.001)

*Asterisks indicate variables with negative beta weights that were positively correlated with the criterion